

How to build models for government: criteria driving model acceptance in policymaking

Daniel Antony Kolkman¹ · Paolo Campo¹ · Tina Balke-Visser¹ · Nigel Gilbert¹

Published online: 3 May 2016

© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Models are used to inform policymaking and underpin large amounts of government expenditure. Several authors have observed a discrepancy between the actual and potential use of models in government. While there have been several studies investigating model acceptance in government, it remains unclear under what conditions models are accepted. In this paper, we address the question “What criteria affect model acceptance in policymaking?”, the answer to which will contribute to the wider understanding of model use in government. We employ a thematic coding approach to identify the acceptance criteria for the eight models in our sample. Subsequently, we compare our findings with existing literature and use qualitative comparative analysis to explore what configurations of the criteria are observed in instances of model acceptance. We conclude that model acceptance is affected by a combination of the model’s characteristics, the supporting infrastructure and organizational factors.

Keywords Decision support systems · Model acceptance · Evidence-based policy · Model development

Introduction

Papers that present a model of some kind often claim that the model could be used to inform policymaking. Yet, only some of those studies present evidence of the actual use of the model in question (van Ittersum and Sterk 2015). This suggests that not all models are as useful to inform policymaking as their authors claim and warrants further investigation into what determines whether a model is used in practice.

✉ Daniel Antony Kolkman
d.kolkman@surrey.ac.uk

¹ Centre for Research in Social Simulation, University of Surrey, Guildford GU2 7XH, Surrey, UK

The subject of model acceptance is particularly relevant in the context of policymaking. Nilsson et al. (2008) argue that during the twentieth century, the issues facing policymakers have become increasingly complex. They suggest that dealing with such issues requires policy to be grounded in an evidence base. Models can be used to contribute to the evidence base and are increasingly used to inform policymakers (van Daalen et al. 2002). In effect, models underpin large amounts of government expenditure and motivate decisions that affect people's lives (Scholten 2008; Treasury 2013).

The use of models in policy has been said to fall short of its potential (McIntosh et al. 2008; Turnpenny et al. 2009). Moreover, research on the broader subject of model use is argued to be in its infancy (McIntosh et al. 2007) and some contend that this research field has failed to progress over the last few decades (Syme et al. 2011). McIntosh et al. (2007) claim that the research agenda has tended to focus on the technical aspects of models, but has largely ignored other criteria that influence their use. Previous research has suggested a variety of such criteria that can drive or constrain model acceptance, for instance, user-friendliness (Fildes et al. 2006; van Delden et al. 2011), interpretability (Baesens et al. 2009; van Delden et al. 2011), model performance (Baesens et al. 2009; van Delden et al. 2011), the user interface (McIntosh et al. 2011; van Delden et al. 2011) and the model development process (McIntosh et al. 2005; Stalpers et al. 2009).

In this paper we contribute to the debate on the acceptance of models in policymaking and the use of models in general. We employ an inductive approach to identify the criteria for model acceptance in a sample of eight models. Data on these models were collected in the Netherlands and the UK by means of semi-structured interviews, observations and archival research. The fieldwork was conducted at ministries, agencies and commercial organizations.

The paper is structured as follows: we first present a working definition of models and a description of what is meant by acceptance in the context of this study and by drawing on existing literature. We then explore existing contributions to the subject of model acceptance in policy. This is followed by an identification of the model acceptance criteria that were observed for the sample of eight models, using methods based on grounded theory and thematic coding. We conclude by juxtaposing these criteria with conclusions from previous research and highlight the criteria that those developing models can influence to improve the usefulness of their models.

Context

Evidence-based policymaking and models

Nilsson et al. (2008) argue that over the course of the twentieth century, the issues facing policymaking have become increasingly complex. Head (2010) points out that such complexity strains both scientific rigour and political management. For instance, complexity clashes with the culture of auditing and target setting, as it obscures the very causality, reductionism, predictability and determinism on which the latter is based (Geyer 2012). By drawing on an evidence base, policymakers can focus on what works and avoid the pitfalls of policy driven by ideology or values (Botterill and Hindmoor 2012).

The effects of the move towards evidence-based policymaking have been criticized (e.g. Fischer 2007). In practice, evidence-based policymaking often fails to live up to its promise. Policymakers operate under continual pressure to reach decisions. Their decisions

will reflect not only beliefs about what works but judgments about what is feasible as well as elements of ideological faith, conventional wisdom and habit. Policymakers are also constrained by resource limitations in terms of time and financial inputs (Botterill and Hindmoor 2012). These insights have given rise to what Head (2010) terms a “new realism”—the idea that we cannot expect to construct a policy system that is fuelled by objective research findings alone. Rather, it suggests that a variety of evidence—not just scientific evidence—will inform the policy process.

Despite these criticisms, the last decades have seen a sharp increase in policy analysis activities, which can be considered as one of the main vehicles for the construction of an evidence base (Nilsson et al. 2008). Nilsson et al. (2008) distinguishes three groups of “tools” for policy analysis: (1) Simple tools such as check-lists, questionnaires, impact tables, process steps or similar techniques for assisting expert judgment, (2) formal tools, such as scenario techniques, cost benefit analysis, risk assessment and multi-criteria analysis, which entail several analytical steps corresponding to predefined rules, methods and procedures, and (3) advanced tools that attempt to capture the more dynamic and complex aspects of societal or economic development by performing computer-based modelling, simulation or optimization exercises.

Considering the lack of consensus on the definition of a model, it is pertinent to explicitly define the concept as it used in this paper. Minsky (1965) defined models as “To an observer B, an object A* is a model of an object A to the extent that B can use A* to answer questions that interest him about A”. In this definition, a model is a method of inquiry into a particular target system. However, for the current purpose, a more specific definition of the form of “object A*” is required. In particular, our focus is the computer implementation of algorithms that policymakers can interact with to inform their decision-making. As such, models can be defined as: “A formal representation of an object that those informing policymakers can use to answer questions about that object”. In the context of this paper, such models sit within the third category defined by Nilsson et al. (2008).

McIntosh et al. (2007) assert that many support-tool technologies contain models. More specifically, many decision support systems, planning support systems, and management information systems utilize models as part of a more extensive software package. Syme et al. (2011) also group these technologies under the same heading. Hence, the support-tool technologies that contain models are also considered in this study.

Models can be used in several ways to contribute to an evidence base. This is reflected in a recent classification of models used in government put forward by the UK Treasury (2013). They discern seven types of models based on the tasks that models are used for: (1) policy simulation models, used for impact analysis, (2) forecasting models that assess the future and provide information for financial planning, (3) financial evaluation models that assess the liabilities of future costs, (4) procurement and commercial models used for value for money evaluation and awards of contracts, (5) planning models for formulating actions based on forecasts, (6) science-based models that are used to understand and forecast natural systems, and (7) allocation models that aid the allocation of funding across government organizations.

Research gaps and challenges

Several advantages are ascribed to using models to inform policy (McIntosh et al. 2007). For instance, van Daalen et al. (2002) suggest that models can serve as eye-openers by placing issues on the political agenda, challenge existing world views, aid consensus

forming and support the management of a particular target system. (Jakeman and Letcher 2003) point out that models provide a way of exploring and explaining trade offs, a tool for adoption and adaptation by stakeholders, a longer term memory of the project methods and a library of integrated data sets.

Despite these advantages and the increase of model use in policy (van Daalen et al. 2002), model use has been argued to fall short of its potential (McIntosh et al. 2008; Turnpenny et al. 2009). Several explanations for this gap have been put forward. For instance, McIntosh et al. (2008) concluded that it is partly a result of the different perceptions of model users and model developers on what a model should look like. Van Delden et al. (2011) suggested a lack of transparency, inflexibility and a focus on technical capabilities as significant impediments to the acceptance of models in policy. Happe and Balmann (2008) implied that models do not necessarily fit well within the day to day work routines of those informing policy.

There is also a lack of clarity in regard to what is meant by model acceptance. McIntosh et al. (2011) distinguish four levels of acceptance: (1) model development has been completed and presented to its intended users, (2) the users have been trained in the use of a model, but there is limited evidence of actual use, (3) the model has been used on a one-off basis, and (4) the model is used routinely as a recognized part of a user's occupation. In this study, we consider the fourth level as model acceptance.

There is no consensus on the extent to which the characteristics of a model affect its acceptance. For example, while Roosenschoon et al. (2012) claim that the acceptance of models depends on the type of model and claim that user requirements for models defy generalization, McIntosh et al. (2011) argue that there is no evidence to support the idea that model characteristics determine model acceptance. They stress, however, the importance of the model development process. Despite these criticisms, we will now look at what criteria have been associated with model acceptance by previous research.

Drivers and constraints to model acceptance in the literature

Vonk and Geertman (2008) suggested that users prefer simple models over advanced ones, user-friendliness is of importance, organizational support for implementing models is often limited, managers consider the implementation of models to be risky, users have limited experience with models, and there is little proof of the added value of models. Furthermore, the professional training of users in the application of models remains limited. This means that they are unfamiliar with the use of such technologies and less likely to adopt them. In addition, the failure of previous technologies to live up to expectations has made users reluctant to use models (Vonk et al. 2007).

Van Delden et al. (2011) suggested that it is necessary for users to interpret model outputs correctly. Also, a model needs to connect to the policy context and provide added value to those working with it by having variables and outputs that relate to the policy context and are meaningful to the users. Another obstacle to model acceptance is that of unrealistic expectations. Model developers should manage the expectations of users by communicating both the potential and the limitations of models to avoid disappointment. The authors argue that “champions” who advocate a model facilitate its adoption and use. They also suggest that the way models handle uncertainty is of importance to the users. Other determinants include data-availability and model runtime.

McIntosh et al. (2011) also argued that the model development process is relevant. During this process, the model developers should not oversell the model and should focus on minimizing costs and training needs. Furthermore, the user interface of the model

should be simple and user-friendly. Technical performance of models is not the only relevant criterion for successful deployment. Increasingly, interpretability is considered to be of importance. It should also be noted that aiming for interpretability sometimes comes at the cost of model performance, and this trade-off between model readability and performance needs to be taken into account (Baesens et al. 2009). Goodwin et al. (2007) proposed that a model should be acceptable to users, be easy to use, offer a flexible range of methods, be viable for commercial exploitation, and facilitate the appropriate mix of judgment and statistical methods.

Methods

Sampling strategy

To investigate the drivers and constraints of model acceptance in policy, we conducted fieldwork on eight different models that are currently in use (see Table 1). Three case studies were conducted in the Netherlands and five in the UK. The cases were selected on the basis of maximum variation sampling. This sampling strategy consists of including as much diversity in the cases as possible. It generally results in a heterogeneous set of cases, which allows the researcher to identify commonalities that cut across the variations (Patton 2002). If such commonalities exist amongst divergent cases, they provide a basis for theory formation.

The maximum variation sampling of cases was guided by a set of selection criteria. The first selection criterion is that of novelty, that is, the number of years since it was first completed. If a model has been around for a long time, those working with it may have developed familiarity with it, affecting its perceived user-friendliness and interpretability. This idea is supported by findings on the adoption of information technology in general (Venkatesh and Bala 2008).

The second selection criterion is model technique. Model characteristics such as performance and the user interface are likely to be related to the technique used in the model. For instance, a particular modelling technique might be computationally intensive or limit the options for an interface. Moreover, Boulanger and Bréchet (2005) demonstrate that different modelling techniques have different strengths and weaknesses for use in policy. Although their work does not claim a link between these strengths and weaknesses and the way in which models are used, it does support the connection between model technique and other model characteristics.

Besides model characteristics, the context in which models are used for policy has been argued to impact how a model is used. However, there has been little work that provides a typification of the contexts of model use, and the relative importance of different aspects of these contexts has not been established. To permit sampling for maximum variation given these limited theoretical foundations, multiple selection criteria were used. By aiming for variation on three selection criteria, some level of variation in terms of contexts could be achieved. The first criterion for the context of model use is model purpose, classified according to the United Kingdom Treasury's seven purposes (Treasury 2013).

The second and third contextual selection criteria are based on the assumption that model use is subject to cultural influences. Diez and McIntosh (2011) put forward this idea and suggest that model use can differ across geographical contexts. For this study, models were sampled from two countries: the UK and the Netherlands. Sampling was limited to these two locations for practical reasons. Following the premise that model acceptance is

Table 1 Case descriptions

Case ID	Model name	# Of interviews	# Of documents	Description
A	Pensim2	8	10	A dynamic micro-simulation model that estimates the future distribution of pensioner incomes. It is developed and used at the Department for Work and Pensions in the UK. This department is responsible for welfare and pensions policy. This model dates back ten years and has an estimated user-base of about twenty people
B	SAFFIER II	9	7	A macro-economic model developed and used at the Dutch Bureau for Economic Policy Analysis. This independent organization delivers economic analysis and forecasts for the government of the Netherlands. SAFFIERII is part of a family of models with a history of several decades. This particular version has been in use for about fifteen years, the model has three users and upwards of 20 end-users
C	Forecasting model ^a	5	8	A linear regression model used to estimate workload at the Dutch Land Registry. The Dutch Land Registry is responsible for the registration of the ownership and location of plots of land. The forecasting model predicts the number of conveyance deeds, mortgages and survey requests. It has been developed by the OTB research institute of the University of Delft, the Netherlands. The Forecasting model is used by about five people
D	EUPA model ^b	5	7	An accounting-type model of energy using products developed and used for the UK Department for Environment, Food & Rural Affairs by ICF International. The EUPA model serves to support the design of policies to improve the resource efficiency of the UK's economy. The model has an estimated user-base of ten people. The Department for Environment, Food and Rural Affairs is responsible for environmental protection, agriculture, fisheries and rural communities in the UK, while ICF International is a consultancy
E	2050 calculator	2	2	An accounting-type model of carbon use developed by the UK Department for Energy and Climate change for internal and external use. The department is responsible for energy security, affordable energy supplies and climate change mitigation in the UK. The 2050 calculator was developed by a team of about five model developers and is accessible to the general public
F	Retail Risk Index	3	3	A model of retailer default risk, developed by a Dutch joint venture between the data broker Locatus and the consultancy Insights4action for use by institutional investors, retailers and government. The model has an estimated user group of about 15 people
G	UKTimes	3	3	An optimization model developed by the University College London Energy Institute for use at the UK Department for Energy and Climate Change. The model was under development at the time of this study

Table 1 continued

Case ID	Model name	# Of interviews	# Of documents	Description
H	Quitsim	2	1	An agent-based model developed by Sandtable for use at Public Health England. Public Health England is a government agency that is responsible for the safeguarding of health and well-being the UK. The model is used to estimate the impact of marketing campaigns on smoking cessation. Its user-base consists of about ten people

^a Author's translation from original Dutch ("Ramingsmodel")

^b A fictitious name was used at the request of the participants

affected by norms and institutions, the host organization was used as a third selection criterion for context.

The two dimensions and the five selection criteria they encompass are a means towards accomplishing maximum variation in the sample of models for this study. However, there exists no exhaustive inventory of all models that are used to inform policy. In more technical terms, this constitutes the absence of a sampling frame to which the selection criteria can be applied.

To prevent sampling bias and to allow for the application of the sampling criteria, a tentative sampling frame was created by combining web-searches, online archives of models, such as the LIAISE Toolkit (LIAISE 2014), and the annex to the UK government review of analytical models (Treasury 2013). By combining these archives and manual web-searches, an overview of 660 models was created. Still, some bias in the sampling could not be prevented as more controversial models might not be publicly advertised. Twenty models were targeted for inclusion in this study, but access was only attained to eight.

Over a period of 2 years, data were collected on these eight models in the form of interviews, documents and observations. The data included 34 semi-structured interviews with model developers and policy analysts. To ensure interview comparability, an interview guide was used. The interview data were supplemented by an archival study of 41 documents such as minutes, model documentation and green papers that reference the model. In addition, in situ observations of model developers and policy analysts were conducted. Taken together, these three methods allowed for validation of the findings through triangulation.

Data analysis

For the analysis of the collected data, we employed a format that is based on the grounded theory paradigm put forward by Strauss and Corbin (1998). The approach consists of *open coding* each interview and attributing descriptors to fragments of texts. In a subsequent stage, these open codes are refined and recombined into more clearly defined concepts. These concepts are then used to recode the interviews and other data for each case. An overview of all the extracted codes was written up in a report form for each case study. The final stage of analysis consisted of merging the codes that were set out in the respective case studies. During this phase, the codes with a direct relation to model acceptance were extracted and a list of text excerpts was created for each observed criterion of model acceptance.

Observed model acceptance criteria

In this section, we identify and define the criteria that contributed to the acceptance of each model. The grounded theory-based process of coding resulted in the eleven criteria of model acceptance in government shown in Table 2. This table also demonstrates which criteria were observed in each particular case. A limitation of such binary categorization is the reduction of nuances or contrasts that may exist in the data. To ameliorate this and to structure the categorization, we employed a simple heuristic. A criterion was considered to be “observed” and assigned a value of “1” if the code for that criterion occurred in two separate sources for the same case. The eleven criteria that emerged from the data can be divided into three groups: model characteristics, organizational characteristics and supporting infrastructure. Two criteria could not be aggregated: reputation and participation in development.

Model characteristics

Model characteristics are attributes of a particular model. Of the eleven criteria, quality, tractability, efficiency and flexibility belong in this category.

Quality

We define quality as the degree in which the model is perceived to be valid. In all eight cases validation was reported to be a key consideration in driving acceptance of the model. For the Forecasting model, concerns about the validity of the previous model motivated the organization to seek development of a new version. For Quitsim, validation was prioritized in the development process in order to create wider support for the model. In the SAF-FIERII case, model results were monitored and validated on a rolling basis in order to ensure the quality of the outcomes.

The quality of the underlying data played an important role in all of the cases. In the Forecasting model case, the participants noted that it was important for the data to originate

Table 2 Observed criteria of model acceptance in policy

Models (see Table 1)	A	B	C	D	E	F	G	H
Reputation	0	1	1	1	0	1	1	1
Participation in development	1	1	1	1	0	0	1	1
Model characteristics								
Quality	1	1	1	1	1	1	1	1
Tractability	0	1	0	0	1	1	1	1
Efficiency	1	1	0	1	0	0	0	0
Flexibility	1	1	0	0	0	0	0	0
Supporting infrastructure								
Compatibility	1	1	0	0	1	1	0	0
Transparency	0	1	0	1	1	0	1	0
Consistency	0	1	0	1	0	0	1	1
Organizational factors								
Organizational conditions	0	0	1	0	1	0	0	0
Advocates	0	0	1	0	1	0	0	0

from reputable sources. In the UKTimes and Quitsim cases, interviewees suggested that the stakeholders' validation of the data had contributed to their acceptance of the model.

Tractability

The capacity of users to understand the model was reported to be of significance in five cases. In the SAFFIERII case, the tractability of the model was a key motivation for it to be accepted. The participants pointed out that the model could not be too complex, because otherwise no single person would be able to comprehend it, rendering the model unusable. In the UKTimes case, the complexity of the previous model was noted by participants to be an important motivation in the development and acceptance of the new one. In the Retail Risk Index case, the intuitive tractability of the chosen modelling technique was reported to have contributed to the acceptance of the model. Policymakers did not wish to understand fully how the model worked, but rather wanted to understand it on an abstract level. One participant of the Retail Risk Index case used a metaphor to make a similar suggestion in pointing out that “in order to drive a car, one doesn't need to know how the engine works. It suffices to know that pressing the accelerator pedal will move the car and the steering wheel can be used to move it in a particular direction”. In the case of the 2050 calculator, participants suggested that exposing policymakers to the mechanics of the model could actually adversely affect organizational acceptance because they were sceptical of the simplifications involved.

Efficiency

In the SAFFIERII, EUPA model and Quitsim cases a decrease in the time taken to run a single iteration of the model was reason to revise the model itself or the structure of model use. In the Pensim2 and SAFFIERII cases, participants noted that it was important for model outcomes to be produced in a timely fashion; the model itself needed to produce figures as fast as possible. In the Quitsim case, the way in which the use of the model was organized was perceived as a main impediment. The interactions between policymakers and model analysts were considered to be time consuming and detrimental to the acceptance of the model. In the other five cases, the participants did not suggest model speed was of importance. However, it should be noted that in these cases the run-time of the model was reported to be in the seconds to ten minutes range.

Flexibility

Flexibility is defined as the potential ease with which a model can be adapted to inform new questions. In the Pensim2 case, it was suggested that the capacity to adapt the model to the changing requirements of the policymakers was important. More specifically, the ability to include particular policy measures in the model was considered critical. A similar point was made about the SAFFIERII model.

Supporting infrastructure

These three criteria refer to the existing or required infrastructure for the use of models. An absence of transparency will not render a model useless, but will impact its acceptance. Likewise, compatibility with existing systems and the consistency of model outputs will impact the perception of whether a model can be used with ease.

Compatibility

Compatibility is the degree to which a model is implemented in a programming language or software platform that the user is familiar with. In the SAFFIERII case, interviewees suggested that the programming language the model was developed in corresponded with the one predominantly used at the organization. They argued that this permitted the long-term maintainability of the model.

A similar point about compatibility was raised by participants in the Pensim2, Forecasting model, EUPA model, and 2050 calculator cases. In these cases, the interviewees suggested that the fact that the model was implemented in a particular spreadsheet software aided its acceptance. Because model analysts were familiar with this type of interface, they could easily interact with the model. In the Retail Risk Index case, the model was implemented in an existing proprietary software platform. The participants reported that the fact that they were familiar with this software aided their acceptance of the model. The Quitsim model was also implemented in a programming language that was familiar to its users.

Transparency

The transparency of the model was considered to be of importance in the SAFFIERII, EUPA model, 2050 calculator and UKTimes cases. The interviewees considered transparency to consist of the ability to review the mechanics of the model and its underlying assumptions. In the Pensim2, SAFFIERII, EUPA model and UKTimes cases, participants suggested that transparency did not necessarily mean that all users should understand the mechanics of the model, but rather that experts reviewing the model should not have any difficulty in doing so.

Consistency

Consistency refers to the degree in which the models' outcomes are perceived to align with previous figures. In the SAFFIERII, EUPA model, UKTimes and Quitsim cases, participants pointed out that inconsistency of model results could lead to confusion and scepticism about the validity of the model, thereby adversely affecting the likelihood of its acceptance.

Organizational factors

Two organizational factors were shown to have an impact on model acceptance.

Organizational conditions

In the case of the 2050 calculator model, the participants reported that one of the main reasons for the model being adopted was that the senior policymakers felt that, at that particular time, they could not be attributed blame if implementation of the model failed. If it succeeded, however, they would be able to benefit from this. In the case of the SAFFIERII model, participants suggested that acceptance of the model had to occur in a context where analysis has to be ongoing. Failure to implement the model in a timely fashion would have had adverse consequences in the progress of the policy process.

Advocates

In the Forecasting model case, the participants suggested that the acceptance of the model was spearheaded by an advocate within the organization. This advocate advised in favour of using a model and the organization responded in suit. A similar process was observed by participants of the 2050 calculator. They argued that the support from one renowned individual contributed strongly to the acceptance of the model within the organization.

Other

The two remaining criteria, reputation of the model developer and participation in model development, cannot be grouped under any of the previous headings.

Reputation

Reputation refers to the reputation of the developer of the model. In the SAFFIERII, Forecasting model, EUPA model, Retail Risk Index, UKTimes and Quitsim cases, development of the model was outsourced to an external party. The participants reported that the reputation of the external developers was of benefit to the acceptance of the model. In the Quitsim case, the participants suggested that the developer first needed to establish its reputation by providing validation of the models results.

Participation in development

With the exception of the 2050 calculator and Retail Risk Index cases, participants suggested that the opportunity to be involved in the process of developing the model had contributed to their acceptance of the model. More specifically, they referred to involvement in defining the mechanisms of a model on a conceptual level. In the Retail Risk Index case, the model developers suggested that the fact that users were not involved in the development process negatively affected their initial willingness to accept it. They argued that the lack of participation might have resulted in a degree of scepticism towards the model.

Relative importance of the criteria

Based on Table 2, the quality of the model was the most observed criterion for acceptance. A further analysis of the relative importance of the eleven criteria was done with the help of Boolean algebra. Analysis techniques based upon such binary logic permit investigation of “conjunctural causation”; the idea that different configurations of criteria may lead to the same result (Ragin 2006). Using the procedure put forward by Schneider and Wagemann (2010), we conducted a crisp set qualitative comparative analysis (QCA) on both the eleven criteria and the grouped criteria.

QCA permits inquiry into the configurations of criteria which in themselves may not be sufficient or necessary in producing a particular outcome. In addition, it assumes that different combinations of criteria may lead the same outcome. Within QCA, a criterion or a set of criteria is defined as necessary if it must be present for a particular outcome to occur. Likewise, the criterion or set is defined as sufficient if by itself it can produce a certain

outcome (Rihoux 2016). QCA is relevant because it allows for exploration of the phase space of possible criteria configurations, which is 2^{11} for the ungrouped set of criteria and 2^5 for the grouped set. QCA can be used to test existing assumptions about the relative importance of these criteria, but also help to develop new theoretical arguments pertaining model acceptance.

As a first step, we considered which combinations of the complete set of eleven criteria were observed where the model was accepted. This analysis did not produce new insights about the relative importance of the separate criteria. However, the analysis on the grouped criteria did offer grounds for further theorizing. An overview of these groupings with their associated observed frequencies can be found in Table 3.

Table 4 shows the outcome of a QCA on the frequency table of grouped criteria of model acceptance. The table lists four solutions that represent different configurations of the grouped model acceptance criteria. These four configurations are the only possible solutions given the values shown in Table 3; no solutions were rejected. Each solution has associated values for raw coverage and unique coverage. Two evaluation criteria can be used to evaluate the outcome of QCA; coverage and consistency. Coverage can be seen as analogous to R^2 in statistical models; raw coverage is a measure of the proportion of cases that is represented by a particular configuration of criteria. Unique coverage represents the proportion of cases that is not covered by any other solution. Consistency resembles the notion of significance and assesses the degree to which a solution or set of solutions agrees in showing the outcome in question (Ragin 2006).

The four solutions cover all configurations of model acceptance criteria in the sample, and as such they have a combined coverage value of 1. Since all models in the sample were adopted, the consistency score of the combined and individual solutions is also 1. The raw coverage of solutions one and two is 0.625. This indicates that 62.5 % of model acceptance can be explained by the combination of model characteristics, participation in the

Table 3 Grouped criteria of model acceptance in policy

Models (see Table 1)	A	B	C	D	E	F	G	H
Reputation	0	1	1	1	0	1	1	1
Participation in development	1	1	1	1	0	0	1	1
Model characteristics	1	1	1	1	1	1	1	1
Supporting infrastructure	1	1	0	1	1	1	1	1
Organizational factors	0	0	1	0	1	0	0	0

Table 4 Qualitative comparative analysis of model acceptance

Solution terms ^{a,b}	Raw coverage	Unique coverage
1. Model*Part*Infra*~Org	0.625	0.125
2. Model*Rep*Infra*~Org	0.625	0.125
3. Model*~Rep*~Part*Infra*Org	0.125	0.125
4. Model*Rep*Part*~Infra*Org	0.125	0.125

Model model characteristics; *Part* participation in development; *Rep* reputation; *Org* organizational factors; *Infra* supporting infrastructure

^a Boolean operators: * indicates AND, ~ indicates NOT

^b Computed using R version 3.1.3 with package ‘QCA’ version 2.0

development process, supporting infrastructure and the absence of organizational conditions; or by model characteristics, reputation, supporting infrastructure and absence of organizational conditions. Solutions three and four can both explain 12.5 % of the cases.

Similar to Table 2, the analysis in Table 4 shows that model characteristics were a necessary condition in the acceptance of all the eight cases. Supporting infrastructure appears in three of the four solutions and can thus be said to also play an important part. From the analysis, it appears that participation in the modelling process and the reputation of the model developer both contribute equally in explaining model acceptance. In the cases where organizational conditions were reported to be of influence on model acceptance, supporting infrastructure, reputation and participation in development were not always pertinent.

Non-acceptance

The eight models in the sample all constitute instances in which the model was accepted. However, the eleven criteria can also enable inquiry into why a model was not accepted. In order to understand non-acceptance it is important to note that model acceptance does not occur in isolation. Rather, an accepted model may replace an existing model and model users may consider several alternative models. In our sample of eight models, we encountered one occasion where an alternative model was considered, but not accepted. We also found one instance where the accepted model replaced a previously used model. Here, “replaced” refers not to a new version such as the move from Pensim to Pensim2, but to a completely new model.

The Dutch Bureau for Economic Policy Analysis considered replacing the macro-economic model SAFFIER II with a dynamic stochastic general equilibrium model (DSGE). A DSGE model was developed by the Dutch Bureau for Economic Policy Analysis because they argued that this type of model was at the forefront of current academic thinking. After the DSGE was completed, it was not accepted because of intractability, inflexibility and inconsistency. The DSGE model was considered to be too inflexible to cope with the variety of questions that the Dutch Bureau for Economic Policy Analysis is asked to answer. More specifically, the Dutch Bureau for Economic Policy Analysis does both projections and scenario analysis. In order to facilitate both, separate DSGE models would need to be developed. This was perceived as a problem because tests showed that separate DSGE models would show inconsistencies in their outputs. Moreover, the DSGE model was considered to be quite difficult to explain to non-expert policymakers.

Before Quitsim, Public Health England (PHE) used a different model to estimate the effectiveness of their smoking cessation campaigns. The quality of this model was put into question by PHE because it seemed to overestimate the impact of their campaigns. Moreover, it was considered to be too inflexible to accommodate the switch to a new marketing strategy. In response, PHE initiated the development of Quitsim to replace the existing model.

Link with existing literature

Of the eleven criteria identified, six have been associated with model acceptance in previous studies on model use: Participation in the development process (McIntosh et al. 2005; Stalpers et al. 2009; McIntosh et al. 2011), tractability (Baensens et al. 2009; van Delden et al. 2011), efficiency, (Baensens et al. 2009; van Delden et al. 2011), the presence

of advocates (van Delden et al. 2011), flexibility (van Delden et al. 2011) and transparency (van Delden et al. 2011).

Although the way models cope with uncertainty was pointed out as a criterion (van Delden et al. 2011), we found little evidence of this in the eight case studies. The information systems literature on technology acceptance offers several theories that have been shown to be predictive of information technology acceptance. The criteria we have identified for model acceptance overlap with two of the most frequently used theories in this field: the Diffusion of Innovations Theory and the Technology Acceptance Model. The Diffusion of Innovations Theory covers the criteria of participation in development, tractability and efficiency. The Technology Acceptance Model suggests that quality, participation in development, tractability, efficiency and the presence of advocates are important for the acceptance of technologies in general.

Towards an understanding of model acceptance

We have demonstrated that a set of eleven criteria can be associated with model acceptance in the sample of eight models. In contrast to Roosenschoon et al. (2012) we find that user requirements for models do not defy generalization. In addition, our findings contradict earlier accounts (McIntosh et al. 2011) that argue that model characteristics do not have an effect on their acceptance.

The focus of this paper was specifically on the identification of criteria that contribute to model acceptance in government. Although it provides several useful insights, the extraction of criteria through thematic analysis has some limitations. Given the modest theoretical foundations and limited extent of existing empirical work, maximum variation sampling was used. In contrast to an alternative sampling strategy in which some variables are tightly controlled, it allowed us to freely identify and explore a variety of criteria that may affect model acceptance. However, this comes at the cost of sacrificing some detail and as such our findings offer limited grounds to establish which criteria were most central to model acceptance. Subsequent research could elucidate the relative importance of the criteria found here by, for instance, focussing on the acceptance or non-acceptance of one model in different contexts.

Moreover, the approach taken does not permit investigation of the processes that contribute to acceptance. As suggested by van Ittersum and Sterk (2015), the acceptance of models is contingent on a social process of learning that takes time. Future work could build on the criteria presented here and consider how models become embedded in the networks of people who make and use them and the practices that facilitate this embedding. Such work could employ frameworks more suited for analysis of dynamic social processes, such as the institutional analysis and development framework (Ostrom et al. 1994), or the repositories of actor-network theory (Mol 2013). A related avenue for future work could consider the similarities and contrasts that exist between different model users' perceptions of a model and how such conflicts are resolved.

Despite these limitations, the proposed set of eleven criteria can provide a useful starting point for developers hoping to improve the likelihood of their models being accepted in policy. Model developers cannot influence all of the criteria that matter to model acceptance in government. They have little ability to influence organizational conditions and the presence of advocates within the organization. Model developers have some, but limited influence over their reputation and the consistency of their model

outcomes with established models. This leaves those developing models with six criteria that they can affect and that they should consider carefully if they are seeking model acceptance in government. The six are quality, tractability, efficiency, flexibility, compatibility and participation in development. These six criteria are not typically addressed in publications that present a model and claim usefulness of that model to policymaking. This suggests that the usefulness of models to policymaking cannot be understood in isolation from the social context in which they are used.

Four of these six criteria (quality, tractability, efficiency and flexibility) are based on the perceptions of the intended users. It is important to note that perception seems to be central to model acceptance. While the quality of a model may be established beyond doubt in an academic community, those involved in policy making may not necessarily perceive it equally favourably. These four criteria could be satisfied by involving the client organization in the model development process and investing in model transparency, for instance, by ensuring the availability of model documentation that is written to the appropriate level of expertise and organizing regular opportunities for model developer and user to meet.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Baesens, B., Mues, C., Martens, D., & Vanthienen, J. (2009). 50 years of data mining and OR: Upcoming trends and challenges. *Journal of the Operational Research Society*, 60, S16–S23. doi:10.1057/jors.2008.171.
- Botterill, L., & Hindmoor, A. (2012). Turtles all the way down: Bounded rationality in an evidence-based age. *Policy Studies*, 33(5), 367–379.
- Boulanger, P. P.-M., & Bréchet, T. (2005). Models for policy-making in sustainable development: The state of the art and perspectives for research. *Ecological Economics*, 55(3), 337–350. doi:10.1016/j.ecolecon.2005.07.033.
- Diez, E., & McIntosh, B. S. (2011). Organisational drivers for, constraints on and impacts of decision and information support tool use in desertification policy and management. *Environmental Modelling and Software*, 26(3), 317–327. doi:10.1016/j.envsoft.2010.04.003.
- Fildes, R., Goodwin, P., & Lawrence, M. (2006). The design features of forecasting support systems and their effectiveness. *Decision Support Systems*, 42(1), 351–361. doi:10.1016/j.dss.2005.01.003.
- Fischer, F. (2007). Deliberative policy analysis as practical reason: Integrating empirical and normative arguments. In F. Fischer, G. J. Miller, & M. S. Sidney (Eds.), *Handbook of public policy analysis: Theory, politics and methods* (pp. 223–236). New York: CRC Press.
- Geyer, R. (2012). Can complexity move UK policy beyond “Evidence-based Policy Making” and the “Audit Culture”? applying a “Complexity Cascade” to education and health policy. *Political Studies*, 60(1), 20–43. doi:10.1111/j.1467-9248.2011.00903.x.
- Goodwin, P., Fildes, R., Lawrence, M., & Nikolopoulos, K. (2007). The process of using a forecasting support system. *International Journal of Forecasting*, 23(3), 391–404. doi:10.1016/j.ijforecast.2007.05.016.
- Happe, K., & Balmann, A. (2008). Doing policy in the lab! Options for the future use of model-based policy analysis for complex decision-making. *107th Seminar of European Association of Agricultural Economists*, 1–13.
- Head, B. W. (2010). Reconsidering evidence-based policy: Key issues and challenges. *Policy and Society*, 29(2), 77–94. doi:10.1016/j.polsoc.2010.03.001.
- Jakeman, A., & Letcher, R. (2003). Integrated assessment and modelling: Features, principles and examples for catchment management. *Environmental Modelling and Software*, 18(6), 491–501. doi:10.1016/S1364-8152(03)00024-0.
- LIAISE (2014). LIAISE Toolkit. <http://www.liaise-kit.eu>. Accessed 15 Feb 2016.
- McIntosh, B. S., Alexandrov, G., Matthews, K., Mysiak, J., & van Ittersum, M. (2011). Preface: Thematic issue on the assessment and evaluation of environmental models and software. *Environmental Modelling and Software*, 26(3), 245–246. doi:10.1016/j.envsoft.2010.08.008.

- McIntosh, B. S., Giupponi, C., Voinov, A. A., Smith, C., Matthews, K. B., Monticino, M., et al. (2008). Bridging the gaps between design and use: Developing tools to support environmental management and policy. In A. J. Jakeman, A. A. Voinov, A. E. Rizzoli, & S. H. Chen (Eds.), *Environmental Modelling, Software and Decision Support: State of the art and new perspective*. Amsterdam: Elsevier.
- McIntosh, B. S., Jeffrey, P., Lemon, M., & Winder, N. (2005). On the design of computer-based models for integrated environmental science. *Environmental Management*, 35(6), 741–752. doi:[10.1007/s00267-004-0172-0](https://doi.org/10.1007/s00267-004-0172-0).
- McIntosh, B., Seaton, R. A. F., & Jeffrey, P. (2007). Tools to think with? Towards understanding the use of computer-based support tools in policy relevant research. *Environmental Modelling and Software*, 22(5), 640–648. doi:[10.1016/j.envsoft.2005.12.015](https://doi.org/10.1016/j.envsoft.2005.12.015).
- Minsky, M. L. (1965). Matter, mind, and models. In *Proceedings of international federation of information processing congress* (Vol. 1, pp. 45–49).
- Mol, A. (2013). Actor-Network Theory: Sensitive terms and enduring tensions. doi:[10.1177/1745691612459060](https://doi.org/10.1177/1745691612459060).
- Nilsson, M., Jordan, A., Turnpenny, J., Hertin, J., Nykvist, B., & Russel, D. (2008). The use and non-use of policy appraisal tools in public policy making: An analysis of three European countries and the European Union. *Policy Sciences*, 41(4), 335–355. doi:[10.1007/s11077-008-9071-1](https://doi.org/10.1007/s11077-008-9071-1).
- Ostrom, E., Gardner, R., & Walker, J. (1994). *Rules, games, and common-pool resources*. Ann Arbor: University of Michigan Press.
- Patton, M. Q. (2002). *Qualitative evaluation and research methods. Qualitative Inquiry*, 3. Thousand Oaks: Sage.
- Ragin, C. C. (2006). Set relations in social research: Evaluating their consistency and coverage. *Political Analysis*, 14(3), 291–310.
- Rihoux, B. (2016). Qualitative comparative analysis (QCA) and related systematic comparative methods: Recent advances and remaining challenges for social science research. *International Sociology*, 21, 679–706.
- Roosenschoon, O., Reis, S., Turnpenny, J., Adele, C., Wascher, D., Weiland, S., et al. (2012). Bridging the gap between modellers and model users, why does this gap exist and what can we do about it? In *2012 International congress on Environmental Modelling and Software*.
- Schneider, C. Q., & Wagemann, C. (2010). Standards of good practice in qualitative comparative analysis (QCA) and fuzzy-sets. *Comparative Sociology*, 9(3), 397–418.
- Scholten, H. (2008). Better modelling practice: An ontological perspective on multidisciplinary, model-based problem solving. PhD thesis, Wageningen University, the Netherlands. Retrieved from: <http://www.narcis.nl/vpub/RecordID/ReM:oi:library.wur.nl:wurpubs%2F372305/Language/nl>. Accessed 10 Dec 2015.
- Stalpers, S. I. P., van Ierland, E. C., & Kroeze, C. (2009). Reconciling model results with user needs to improve climate policy. *Environmental Science & Policy*, 12(7), 959–969.
- Strauss, A., & Corbin, J. (1998). *Techniques and procedures for developing grounded theory*. Thousand Oaks: Sage Publications Inc.
- Syme, G., Bennett, D., Macpherson, D., & Thomas, J. (2011). Guidelines for policy modellers—30 years on: New tricks or old dogs? In *19th International congress on modelling and simulation* (pp. 12–16).
- Treasury, H. M. (2013). *Review of quality assurance of Government analytical models: Final Report*.
- Turnpenny, J., Radaelli, C. M., Jordan, A., & Jacob, K. (2009). The policy and politics of policy appraisal: Emerging trends and new directions. *Journal of European Public Policy*, 16(4), 640–653. doi:[10.1080/13501760902872783](https://doi.org/10.1080/13501760902872783).
- van Daalen, C. E., Dresen, L., & Janssen, M. A. (2002). The roles of computer models in the environmental policy life cycle. *Environmental Science & Policy*, 238, 1–11.
- van Delden, H., Seppelt, R., White, R., & Jakeman, A. J. (2011). A methodology for the design and development of integrated models for policy support. *Environmental Modelling and Software*, 26(3), 266–279. doi:[10.1016/j.envsoft.2010.03.021](https://doi.org/10.1016/j.envsoft.2010.03.021).
- van Ittersum, M. K., & Sterk, B. (2015). Computerized models: tools for assessing the future of complex systems? *The Tools of Policy Formulation: Actors, Capacities, Venues and Effects*,. doi:[10.4337/9781783477043.00016](https://doi.org/10.4337/9781783477043.00016).
- Venkatesh, V., & Bala, H. (2008). Technology acceptance Model 3 and a research agenda on interventions. *Decision Sciences*, 39(2), 273–315. doi:[10.1111/j.1540-5915.2008.00192.x](https://doi.org/10.1111/j.1540-5915.2008.00192.x).
- Vonk, G., & Geertman, S. (2008). Improving the adoption and use of planning support systems in practice. *Applied Spatial Analysis and Policy*, 1(3), 153–173. doi:[10.1007/s12061-008-9011-7](https://doi.org/10.1007/s12061-008-9011-7).
- Vonk, G., Geertman, S., & Schot, P. (2007). A SWOT analysis of planning support systems. *Environment and Planning A*, 39(7), 1699–1714. doi:[10.1068/a38262](https://doi.org/10.1068/a38262).